

# Chendi Li

lichendi.cs@gmail.com

PhD student | High-Performance Computing

Personal website:  
<https://www.lichendi.top>

I am currently a PhD student at the University of Utah, and supervised by Prof. **P. (Saday) Sadayappan**. I got my master's degree from the State Key Laboratory of Computer Architecture, Institute of Computing Technology, Chinese Academy of Sciences, and supervised by Prof. **Yunquan Zhang**. My research interests include high-performance computing, matrix/tensor Optimization, and high-performance machine learning.

## RESEARCH INTERESTS

- High-Performance Computing
- Matrix/Tensor Optimization
- High-Performance Machine learning

## EDUCATION

PhD student in Computer Science, University of Utah	Sep 2022 — expected June 2027
Master in Computer Science, Institute of Computing Technology, Chinese Academy of Sciences	Sep 2019 — July 2022
Bachelor in Computer Science, Hunan Agricultural University	Sep 2014 — June 2018

## RESEARCH EXPERIENCES

Graduate Student Research Assistant University of Utah	Sep 2022 — Present
Graduate Student Research Assistant State Key Laboratory of Computer Architecture, Institute of Computing Technology, Chinese Academy of Sciences	Sep 2019 — July 2022
Undergraduate Research Assistant State Key Laboratory of Computer Architecture, Institute of Computing Technology, Chinese Academy of Sciences	Jan 2018 — June 2019

## PUBLICATIONS

- [ICS 2024] **Chendi Li**, Yufan Xu, Sina Mahdipour Saravani, Saday Sadayappan, Accelerated Auto-Tuning of GPU Kernels for Tensor Computations
- [TPDS 2024] Cunyang Wei, Haipeng Jia, Yunquan Zhang, Jianyu Yao, **Chendi Li**, Wenxuan Cao. IrGEMM: An Input-Aware Tuning Framework for Irregular GEMM on ARM and X86 CPUs.
- [ICS 2023] T Chen, H Jia, Y Zhang, K Li, Z Li, X Zhao, J Yao, **Chendi Li**, et al. OpenFFT: An Adaptive Tuning Framework for 3D FFT on ARM Multicore CPUs
- [ISPA 2021] **Chendi Li**, Haipeng Jia, Hang Cao, et al. AutoTSMM: An Auto-tuning Framework for Building High-Performance Tall-and-Skinny Matrix-Matrix Multiplication on CPUs
- [ICPADS 2021] Jianyu Yao, Boqian Shi, Chunyang Xiang, Haipeng Jia, **Chendi Li**, et al. IAAT: An Input-Aware Adaptive Tuning framework for Small GEMM
- [HPCC 2021] Tun Chen, Haipeng Jia, Zhihao Li, **Chendi Li**, et al. A Transpose-free Three-dimensional FFT Algorithm on ARM CPUs
- [HPC China 2020] **Chendi Li**, Guangting Zhang, Haipeng Jia. Fast Computation of Elementary Functions on ARM Platforms

## RESEARCH PROJECTS

Apache TVM, Contributor	Feb 2023 — Feb 2024
• Designed and developed the dynamic gradient algorithm in TVM. Experimental evaluation on a number of matrix-matrix multiplication and 2D convolution kernels demonstrates an order-of-magnitude improvement in auto-tuning time to achieve the same level of code performance. The paper was accepted by ACM ICS 2024.	

**AutoTSM, Author****Nov 2020 — June 2022**

- Designed and developed AutoTSM, which is used to build high-performance tall-and-skinny matrix multiplication on mainstream CPUs. AutoTSM can speed up convolution layers in real-world deep learning applications, and the performance is competitive with Intel OneMKL and outperforms all conventional GEMM implementations. This work was published in IEEE ISPA 2021 and TPDS'24.

**OpenBLAS, Contributor****Nov 2020 — June 2022**

- Optimized pre-pack matrix-matrix multiplication and triangular solve with multiple right-hand-sides (TRSM) on ARMv8 and X86 platforms. OpenBLAS is one of the most famous open-source BLAS libraries.

**IAAT, Contributor****Nov 2020 — Feb 2022**

- Launched the project and investigated JIT tools for small GEMM. IAAT is a template-driven just-in-time (JIT) small GEMM framework targeting CPUs. This work was accepted by IEEE ICPADS 2021.

**OpenFFT, Contributor****Jan 2018 — Feb 2021**

- Optimized small-scale FFT, and contributed to multi-threading and 2D-FFT. AutoFFT is a template-based FFT codes auto-generation framework that contributes to many Chinese vendors' libraries. This work was published in SC'19, TPDS'20, HPCC'21, and ICS'23.

**OpenVML, Contributor****Jan 2020 — Oct 2020**

- Enhanced the math functions by manipulating IEEE 754 floating points. OpenVML is a vector mathematical library. It achieves an outstanding performance improvement compared to C standard library and ARMPL. This work was accepted by HPC China 2020.

**AWARDS & HONORS**

---

<b>2022</b>	Ph.D. Student Fellowship
<b>2021</b>	First-class scholarships
<b>2020</b>	Second-class scholarship
<b>2019</b>	Third-class scholarship, Outstanding intern in PerfXLab
<b>2015</b>	Collegiate programming contest first prize; Outstanding volunteer

**TECHNICAL SKILLS**

---

<b>Tools</b>	NVIDIA Nsight, Linux, Git, Vim, CMake, GDB, OpenMP, Pthreads
<b>Programming/Scripting</b>	C, C++, Latex, Python, Assembly, NEON and X86 intrinsic, CUDA